AWARD NUMBER: W81XWH-16-1-0135

TITLE: **Characterization of mTOR-Responsive Truncated mRNAs in Cell Proliferation**

PRINCIPAL INVESTIGATOR: Jeongsik Yong

CONTRACTING ORGANIZATION: Regents of the University of Minnesota
Minneapolis, MN 55455

REPORT DATE: July 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2017	Annual	1 Jul 2016–30 Jun 2017

4. TITLE AND SUBTITLE

Characterization of mTOR-Responsive Truncated mRNAs in Cell Proliferation

5a. CONTRACT NUMBER

5b. GRANT NUMBER
8W1XWH-16-1-0135

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jeongsik Yong

E-Mail:jyong@umn.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Regents of the University of Minnesota, Office of Sponsored Projects Admin, 450 McNamara Alumni Center, 200 Oak St SE, Minneapolis, MN

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Defective Tuberous Sclerosis Complex (TSC) 1 or 2 gene leads to deregulated mTOR activation and consequent cell proliferation/growth. Thus, studying the mTOR pathway at a molecular level is fundamental to understand TSC pathogenesis. We recently discovered genome-wide alterations of polyadenylation site in mRNAs. These findings identify a previously uncharacterized role for mTOR in modulating 3'-UTR length of mRNAs by alternative polyadenylation (APA). Another outcome of APA in the mTOR-activated transcriptome is an early termination of mRNA transcription to produce truncated mRNAs with polyadenylation in upstream introns/exons. Truncated mRNAs contain distinct molecular signatures at both RNA and protein levels: the new 3'-end of mRNAs is from introns (intronic 3'-end) and it generates a brand new C-terminus protein sequence encoded from introns. Thus, it is likely that activation of mTOR adds new molecular signatures to functional transcriptome and proteome by alternating polyandeylation. In this reporting period, we profiled truncated mRNAs and their protein products using RNA-seq, 3'-end seq, high capacity mass spectrometry and bioinformatics tools. We developed a new bioinformatics tool to integrate RNA-seq and 3'-end seq and this tool makes it possible to reveal truncated mRNAs genome-wide. In silico proteogenomics database for intron-coded C-terminus of truncated proteins has been developed and used for search of truncated proteins. Additional new software to find unannoated truncated mRNAs are under development.

15. SUBJECT TERMS
truncated mRNA, C-terminomics, alternative polyadenylation, IntMAP, mTOR, TSC

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	14	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

1. INTRODUCTION

Most, if not all, cancer cells proliferate much faster than normal cells. Thus, studying how cancer cells proliferate faster than normal cells is a key in understanding cancer pathogenesis. The mammalian target of rapamycin (mTOR) pathway is a cellular pathway that controls cell proliferation and this pathway is commonly deregulated in many cancers. In addition, Tuberous Sclerosis Complex (TSC) negatively regulates the mTOR pathway. Therefore, studying the role of mTOR pathway in cell proliferation is important to understand the pathogenic mechanism by TSC. When cells are activated to proliferate, the first thing they do is producing a lot of proteins. To make more proteins in cells, they need to make more messenger RNAs (mRNAs) from DNA. The whole procedure is called gene expression and mRNA is a key molecule in this procedure. Thus, the questions of how mRNAs are made and how they are regulated in cancer mechanisms are important questions to ask to understand cancer at a molecular level. Generally, mRNA undergoes very complicated process to make it competent for protein synthesis in cells. Recently, we discovered a pervasive production of truncated mRNAs when mTOR is activated in cells. The truncated mRNAs are produced by dysregulation of one of the steps during mRNA synthesis in cells. The cellular consequence of this phenomenon is the production of truncated proteins. Usually, fundamental elements of many proteins are consisted of catalytically active domains and regulatory domains. The active domain represents the function of a protein and the regulatory domain is a platform for fine-tuning of the protein activity regulated by other cellular proteins. Interestingly, many truncated proteins produced by mTOR activation were lacking the regulatory or catalytic domain. This suggests that mTOR activation produces many deregulated “super isoform” proteins by truncation and this could be a driver to fast cell proliferation and cancer initiation at a molecular level. Our goals in this proposal are to find them and understand their function in cell proliferation using a series of experiments employing high profiling technologies including next generation sequencing and multi-dimensional targeted LC-MS/MS. More importantly, we will narrow down the list of truncated mRNAs that are crucial for cell proliferation. The identified truncated mRNAs will be new targets in cancer-related research and provide brand new molecules that function as a driver in TSC-related pathogenesis.

2. KEY WORDS

Mammalian target of rapamycin, mTOR, Tuberous Sclerosis Complex, TSC, truncated mRNA, alternative polyadenylation, truncated protein, C-terminomics, 3’-end seq, IntMAP

3. ACOMPLISHMENTS

What were the major goals of the project?

As outlined below, the major goals of the projects during the first funding period were to identify mTOR-responsive truncated mRNA and protein isoforms using the combination of high profiling technologies and bioinformatics tools. The specific aim 1 is divided into two major tasks which will cover the topic of functional transcriptome and proteome. Each major task contains a series of experiments and goals to achieve, which are listed under the subheading of each major task.

Aim 1: Identify mTOR-responsive truncated mRNA and protein isoforms

Major task 1.1: Profiling of mTOR-responsive truncated mRNAs

PacBio Iso-seq (months 1-3): Minor changes in this approach. Cost-matching alternatives to this approach were taken and now completed.

RNA-seq in the presence of Torin1 (months 1-3): Completed.

Assembly of Isoform database and integration of RNA-seq with Iso-seq (months 4-6): With minor modifications, this approach is complete.

Major task 1.2: Development of targeted mass spectrometry for truncated proteins

In silico C-terminal database assembly (months 7-8): Completed.

Targeted Mass spectrometry (months 9-12): Completed.

What was accomplished under these goals?

An integrative method, IntMAP, to profile alternative polyadenylation:

We developed a novel algorithm called IntMAP (*Int*egrative *M*odel for *A*lternative *P*olyadenylation), which integrates RNA-seq and 3'-end-seq data for exhaustive analysis of Alternative cleavage and polyadenylation (APA) events (Figure 1). In IntMAP, first the position of multiple polyadenylation sites in a gene is defined and the 3'-UTR isoforms of the gene are accordingly deduced. Then, the quantitative information of RNA-seq and 3'-end-seq data is integrated to calculate the expression level of inferred 3'-UTR isoforms. Two elements in IntMAP work systemically to help the quantitation of isoform expression. The first element promotes the isoform expression to comply with the observed read counts from RNA-Seq data. The second element encourages the consistency between the isoform expression learned from RNA-Seq and 3'-end-Seq data. After the quantitation by IntMAP, the calculated expression level of different 3'-UTR isoforms is applied to the chi-squared test to determine the APA event of a gene in a biological context (Figure 1).

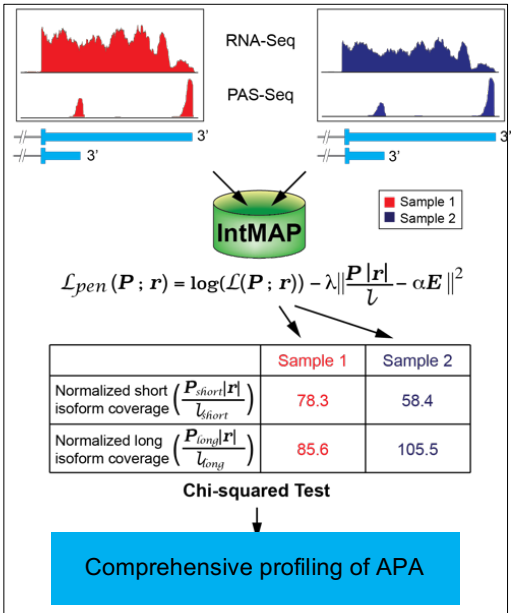

Figure 1. The development of IntMAP software to profile alternative polyadenylation.

Bipartite expression profile of truncated mRNAs upon the cellular mTOR activity:

To identify APA isoforms that could be crucial for cell proliferation, we first profiled truncated mRNAs that are differentially expressed upon the changes in the cellular mTOR activity. Consistent with our findings in WT and Tsc1-/- MEFs, numerous genes increased or decreased their truncated mRNA isoform expression upon the inhibition of mTOR (Figure 2). We consider up-regulated truncated mRNAs in the presence of Torin 1 as the normal tissue-enriched truncated mRNAs and down-regulated truncated mRNAs in the presence of Torin 1 as the TSC-related truncated mRNAs. The profile of differentially expressed truncated mRNAs partially overlapped between two comparison datasets (WT vs. Tsc1$^{-/-}$ MEFs, Tsc1$^{-/-}$ mock vs. Tsc1$^{-/-}$ +Torin 1), suggesting that truncated mRNAs might present additional molecular signatures defining the characteristics of TSC biology.

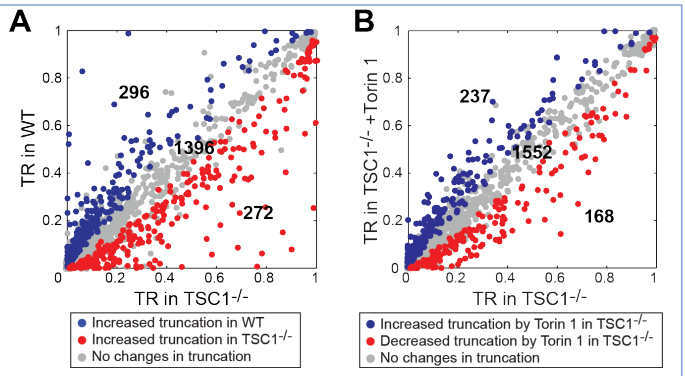

Figure 2. Bipartite expression profile of truncated mRNAs in breast cancer cell lines. MCF7 and BT549 breast cancer cells were cultured in the presence (mimics normal) or absence (mimics tumor) of Torin1. Truncated mRNAs were cataloged using the integration of RNA- and PAS-seq method we developed. Red dots indicate the genes showing truncation in tumor mimicking breast cancer cells. Blue dots indicate the genes showing truncation in normal mimicking breast cancer cells.

<u>Integration of in silico proteogenomics and LC/MS-MS to survey intron-coded peptide sequences:</u>

We developed a new workflow to identify intron-coded peptide sequences in the C-terminal of truncated proteins (Figure 3). The original C-terminomics (from a published protocol) using chemical blocks did not work out in our lab. An SRM approach for targeted LC-MS-MS could identify intron-coded peptides but the efficiency was low (10 proteins identified out of 1151 candidate proteins).

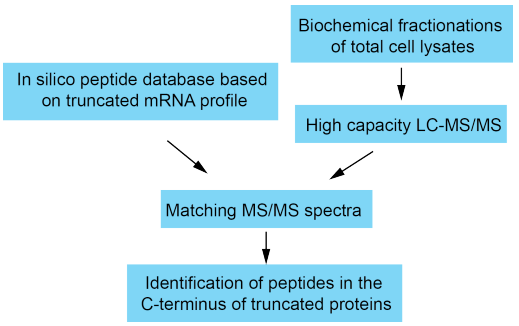

Figure 3. A workflow for C-terminomics

We developed a new method integrating in silico proteogenomics database, heavy biochemical fractionations and high capacity LC-MS/MS to conduct C-terminomics (Figure 3). In this approach, we first conduct biochemical fractionations by size and use each fraction for tryptic digestions and LC-MS/MS experiments. In parallel, in silico peptide sequence database is prepared based on the profile of truncated mRNAs. High capacity LC-MS/MS experiments are performed and the resulting high resolution MS/MS spectra are matched to the peptide database. The matched sequences are considered as C-terminal peptides from truncated proteins. As shown in the table above, we were able to identify 779 matching MS/MS spectra from this approach. They represent 165 unique peptides coming from 107 truncated proteins. Considering 1151 possible truncated proteins, we were able to validate 10% of them. Previous SRM only identified 1% of them.

Total number of peptides in in silico proteogenomics database	Total number of MS/MS spectra matched	Total number of peptides identified	Total number of proteins identified
1151	779	165	107

What opportunities for training and professional development has the project provided?

Nothing to report.

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

Development of IntMAP: The developed software is designed to map the location of polyadenylation sites in a transcriptome. The software will be released as an open source and the user-friendly version of it can be downloaded freely. This will help research communities of human diseases caused by the malfunction of RNA processing.

Gene profile for truncated mRNA and protein isoforms: Our data provide evidence that truncated mRNA and protein isoforms do exist in cells. More importantly, we identified truncated isoforms whose expressions are highly dynamic upon the changes of cellular mTOR activity. There new isoforms relevant to mTOR biology could be important for TSC pathogenesis. The list of these isoforms will help understanding TSC pathogenesis from a different angle since they could be hidden players in TSC biology.

In silico proteogenomics database and the workflow for C-terminomics: We have developed a new workflow for mass spectrometry-based C-terminomics to identify the peptides encoded from intron regions. The database for mouse is now complete and available for the research community. In addition, a systemic approach to identify those peptides using mass spectrometry is also available. These will have a broad impact on research fields from proteomics to genomics to pathobiology.

What was the impact on other disciplines?

Nothing to report.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

Nothing to report.

5. CHANGES/PROBLEMS

Changes in approach and reasons for change

One experimental approach of the major task 1.1, PacBio Iso-seq, was changed to alternative approaches including the 3'-end seq, integration of two sequencing methods and the resulting development of software. The reasons for the change of one experimental approach came from an unexpected increase in the service cost for PacBio Iso-seq. After the award for the project has been decided, a new and improved version of PacBio Iso-seq was only available with a significant increase in the cost of sequencing. The price increased more than twice than originally calculated. Since the cost for RNA-seq experiments also increased with a newer version of technologies, it was impossible to afford both profiling methods to achieve the goals of the specific aim 1. Therefore, approaches including alternative sequencing method along with the development of bioinformatics tool to integrate two sequencing data have been taken.

Actual or anticipated problems or delays and actions or plans to resolve them

An actual problem of increasing sequencing cost happened. To overcome this problem, we conducted 3'-end seq which will catalog transcriptome-wide polyadenylation sites in mRNAs. Then we developed an algorithm that applies the positional information of poly(A) sites from 3'-end seq data to RNA-seq data and quantitate the alternative poly(A) isoform expression. From this approach, we were able to analyze the expression of Torin 1-responsive truncated mRNAs. The 3'-end seq experiments and the development of a new software could bring down the cost of sequencing (thus match the budget) and still achieve the goals suggested in the grant proposal.

Changes that had a significant impact on expenditures

Nothing to report.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Nothing to report.

Significant changes in use or care of human subjects

Nothing to report.

Significant changes in use or care of vertebrate animals.

Nothing to report.

Significant changes in use of biohazards and/or select agents

Nothing to report.

6. PRODUCTS

Publications, conference papers, and presentations

Yeh HS, Zhang W, Yong J. Analyses of alternative polyadenylation: from old school biochemistry to high-throughput technologies. BMB reports. 2017; 50(4):201-207. PMID: 28148393, PMCID: PMC5437964

Website(s) or other Internet site(s)

Nothing to report.

Technologies or techniques

Nothing to report.

Inventions, patent applications, and/or licenses

Nothing to report.

Other Products

Software:

IntMAP (*Int*egrative *M*odel for *A*lternative *P*olyadenylation) to identify alternative polyadenylated mRNAs genome-wide has been developed. The source code for the software will be open to public once the manuscript is published in a scientific journal. In addition, a user-friendly free software package will be prepared for easy download and distribution.

Data:

RNA-seq and 3'-end seq data reported in this annual report will be deposited to public data depository once the manuscript is published.

Database:

In silico proteogenomics database for genome-wide intron-coded peptide sequences will be publicly available once the manuscript is published.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	Jeongsik Yong	Name:	Rui Kuang
Project Role:	PI	Project Role:	Co-I
Researcher Identifier:	N/A	Researcher Identifier:	N/A
Nearest person month worked:	5	Nearest person month worked:	2
Contribution to Project:	Dr. Yong has been leading the entire project and has coordinated collaborations.	Contribution to Project:	Dr. Kuang mentored Dr. Zhang to develop an algorithm and software for the detection of truncated mRNAs.
Funding Support:	NIGMS R01 and DoD TS	Funding Support:	NSF, NIGMS R01 and DoD TS
Name:	Tim Griffin	Name:	Hsin-Sung Yeh
Project Role:	Co-I	Project Role:	Graduate Student
Researcher Identifier:	N/A	Researcher Identifier:	N/A
Nearest person month worked:	2	Nearest person month worked:	3
Contribution to Project:	Dr. Griffin mentored Dr. Park and Mr Yeh for C-terminomics and data interpretations.	Contribution to Project:	Mr. Yeh collaborated with Dr. Park for C-terminomics.
Funding Support:	NSF, NIGMS R01 and DoD TS	Funding Support:	NIGMS R01 and DoD TS
Name:	Meeyeon Park	Name:	Wei Zhang
Project Role:	Research Associate	Project Role:	Research Associate
Researcher Identifier:	N/A	Researcher Identifier:	N/A
Nearest person month worked:	6	Nearest person month worked:	6
Contribution to Project:	Dr. Park collaborated with Dr. Park for C-terminomics.	Contribution to Project:	Dr. Zhang developed an algorithm and software for the detection of truncated mRNAs.

Funding Support:	NIGMS R01 and DoD TS	Funding Support:	NSF, NIGMS R01 and DoD TS
Name:	Jaewoong Chang	Name:	
Project Role:	Research Associate	Project Role:	
Researcher Identifier:	N/A	Researcher Identifier:	
Nearest person month worked:	6	Nearest person month worked:	
Contribution to Project:	Dr. Chang worked on the experimental set up for the profiling of mTOR-responsive truncated mRNAs and the validation of profiling results using wet experiments.	Contribution to Project:	
Funding Support:	NIGMS R01	Funding Support:	

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to report.

What other organizations were involved as partners?

Nothing to report.

8. SPECIAL REPORTING REQUIREMENTS

Nothing to report.

9. APPENDICES

Nothing to report.